INFORMATION AND CONTROL 1, 76–81 (1957)

A Note on a Law of Berry and on Insistence Stress

BENOIT MANDELBROT

Faculté des Sciences, Université de Genève

INTRODUCTION AND SUMMARY

One can distinguish a number of "kinds" of linguistic stress. One is the "tonic stress," which falls upon some syllables of a word, always the same ones. Another, which will be called "insistence stress," may fall, sometimes, on a word as a whole, presumably when it is desired to "insist" on this word. It has been vaguely observed by most speakers that this kind of stress falls more often upon the rare words than upon frequent ones. This observation has been now vindicated by an experimental law, discovered by Berry (1953), which makes it possible to begin to study the insistence stress from the viewpoint of information theory. However, unfortunately but hardly evitably, the data from which this law was drawn are based upon the judgment of some observer or group of observers, and not upon any explicit definition of the stress. We have, despite this difficulty, decided to put a model of Berry's law on record,[1] (a) because it is closely linked with one of our models for Zipf's law for the rank frequency relationship for words; (b) because it is extremely simple, and may therefore help make the definition of stress more precise; (c) because it is the exact opposite of the obvious model for the tonic stress in languages such as French, so that one may hope than one holds two extreme behaviors for stress, between which would be found all the other behaviors, yet to be studied.

Essentially, to derive Berry's law, it is, at least, sufficient to conjecture (a) that a word is stressed when at least one of the ultimate units, into which it can further be decomposed, carries a certain special "mark"; (b) that the occurrence of this mark is ruled by a recurrent random proc-

[1] This model was mentioned in a seminar at Massachusetts Institute of Technolgy in the spring of 1953. It was later reported on, in a paper presented at the Third London Symposium on Information Theory, held in 1955; but it was withdrawn from publication in the Proceedings of that Symposium, and later published by Butterworths, London, and by Academic Press, New York.

ess, essentially identical with the random process that segments the stream of ultimate units into words; (c) that the space process and the stress process are stochastically uncorrelated. Hypothesis (b) also holds for the tonic stress in French, for example, if the ultimate units are taken to be the syllables; but (c) does not hold there since the two processes are then fully determined by each other (there is the fixed one-syllable phase difference between the tonic stress and the end of the word).

STATEMENT OF THE LAW

Suppose that the probability for the occurrence of word W_r (say $p(r)$) and the probability that this word be stressed when it occurs (say $q(r)$) are both well defined. Suppose further that $p(r)$ is a decreasing function, that is, that the index r is defined as being the one which ranks the words in the order decreasing probability. Let $F(r) = \sum_{x=1}^{r} p(x)$. Then, Berry's law states that:

$$q(r) = F(r) = \sum_{x=1}^{r} p(x)$$

and a generalized Berry's law states that:

$$1 - q(r) = P[1 - F(r)]^A$$

where the parameters P and A are no longer both necessarily equal to 1.

Recall also the relationship between $p(r)$ and r, discovered by Estoup and Zipf (see Mandelbrot, 1953, 1955, 1957a, b): $p(r) = Kr^{-B}$. Thus, Berry's law becomes:

$$1 - q(r) = Pr^{-(B-1)A}.$$

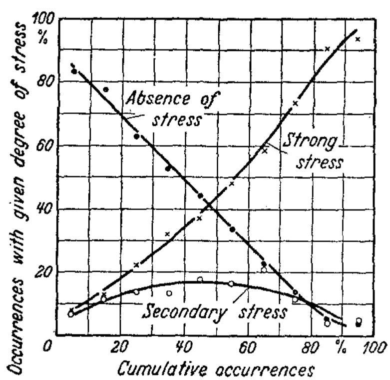

FIG. 1. Percentage of occurrence of words with various degrees of stress as a function of cumulative occurrences when words are arranged in descending order of frequency.

A CONJECTURE

Since the insistence stress applies to whole words, its definition, and the understanding of its structure, are conditional upon the definition and the understanding of the properties of the words themselves. Consider then a sequence of linguistic units, more elementary than the words. In information theory, this is considered to be a random sequence, and the fact that a unit is the last one of a word is a random event on this sequence. Any theory of the formation of words, by "composition" of units, or "segmentation" of the stream of those units, has inevitably two aspects: to explain the facts which have been observed, and to help to define better the concepts to which these facts refer. In particular, an informational theory of natural languages should provide a model for the random process of word ending. In previous publications (Mandelbrot, 1953, 1955), we have shown that, in the first approximation, one may assume that the process of word ending has a very short memory, and, in fact, that it is Markovian: inter-word relationships do not have any influence upon the statistics of the words themselves.

Let us test a similar conjecture in the case of stress, unrealistic as it may seem *a priori*: *assume that the insistence stress* (just like tonic stress, but not necessarily in the same fashion) *is based upon a random event on the stream of units more elementary than the words*; *and* (most strongly) *that this event is Markovian and statistically independent from the event that delimits the words.* Thus the presence of insistence stress would result from the following special kind of interplay between the two processes: *a whole word would be stressed if, and only if, a stress mark occurs on at least one of its elementary units.*

DERIVATION OF BERRY'S LAW, IN THE SIMPLEST CASE

In the simplest case, one of our 1953 theories of segmentation of speech into words assumes that the stream of the elementary units of language is a stream of independent units, of which G are equiprobable, of probability $(1 - p_1)/G$, and the $(G + 1)$st has probability p_1, and occurs at the end of each word, and only there. Because of this asymmetric property, this $(G + 1)$st letter is somewhat "improper"; it will be called "space."

As an index for a word, replace r by the number of words it contains, C (space included); this is possible because, in the present case, $p(r)$ depends upon r through C only:

$$p(C) = p_1(1 - p_1)^{C-1}G^{-(C-1)}$$

$$[\text{with } p(C = 1) = p(r = 1) = p(\text{space}) = p_1]$$

Now the number of different words of length C is

$$S(C) = G^{C-1}$$

Hence:

$$F(C) = \sum_{x=1}^{C} G^{x-1}(1 - p_1)^{x-1}G^{-(x-1)}p_1 = 1 - (1 - p_1)^C.$$

Write now the hypothesis that the probability of a letter carrying a stress mark is p', independent of anything happening to preceding letters. Then the probability of a whole word being stressed still depends upon whether the space can carry this mark. If it can carry it,

$$q(C) = 1 - (1 - p')^C,$$

hence

$$1 - q(C) = [1 - F(C)]^A$$

which is the generalized law of Berry with $P = 1$, and $A = \log(1 - p')/\log(1 - p_1)$. In particular, if $p' = p_1$, $A = 1$. Thus, the nongeneralized law of Berry corresponds to the case that the average recurrence time of "space," $1/p_1$, is also equal to the average recurrence time of the stress mark. It would not be surprising that p_1, which is an important feature of speech since it rules the segmentation into words, would have a much wider significance. If this fact could be confirmed, it would make possible a closer study of the psychological and physiological level by which segmentation is conditioned. Anyway, from Berry's original data, A could not be very different from 1.

If the space cannot carry the stress mark, that is, if this mark on the space does not influence the presence of stress,

$$q(C) = 1 - (1 - p')^{C-1},$$

so that

$$1 - q(C) = P[1 - F(C)]^A,$$

with the same A as above and with $P = 1/(1 - p')$

GENERALIZATION

Let us simply state here that Berry's law may also be derived if the letters were not equally probable, and even if they were not independent from each other but linked by a Markoff process. Note in any case that if, in the preceeding section, p' and p_1 are both small, $A = p'/p_1$. This means that A does not depend explicitly upon the letter process but only on the relative frequencies of the stress mark and of the space; this approximate result still holds in the Markovian case.

CONCLUSION

The property last referred to in the preceding section means that the law of stress is not very critically dependent upon the random process which is postulated for the stress mark. The same independence of properties relative to words, from the properties of letters, has been previously found in the case of our models for the law of Estoup-Zipf. It should also be compared with the relative independence of the properties of large masses of gases, upon the mechanical properties of systems of molecules, postulated to explain those properties. It would seem that Berry's law would still hold for stress with even more general assumptions relative to the stress mark: the essential feature would be the postulation of some kind of stationarity of the stress process, and its comparative independence from the space process.

Because of this feature, the statistical linguistic laws obtained in this fashion apply to *any* language, just as the thermodynamical laws, as obtained by the models of the kinetic theory of gases, applied to *every* chemical compound. This is, of course, because the description of the facts given by the bulk laws is very much less detailed than that provided by the actual laws of the interaction of physical or linguistic molecules. We give elsewhere a detailed discussion of the philosophy of these "macrolinguistic laws" or "linguistic macrostructures," and of their relationship with the usual microscopic laws (see Mandelbrot, 1957a). Note only one peculiarity of macrolinguistic laws, compared to macrophysical ones: the concepts to which they refer are ill-defined, and data are scarce; and the "fortunate unsensitivity" of the models, relative to the assumptions made, which makes a theory possible under such circumstances, means, conversely, that the theory can be of very little help in making the concepts more precise, unless it is further developed.

The empirical as well as theoretical study of this problem seems to be full of promise.

RECEIVED: November 15, 1956

REFERENCES

BERRY, J. (1953). *In* "Communication Theory" (W. Jackson, ed.), p. 392. Academic Press, New York.

MANDELBROT, B. (1953). *In* "Communication Theory" (W. Jackson, ed.), p. 486. Academic Press, New York.

MANDELBROT, B. (1955). *In* "Information Networks" (E. Weber ed.), p. 205. Polytechnic Institute, Brooklyn, New York.

MANDELBROT, B. (1957a). Linguistique statistique macroscopique, I. *In* "Logique, langage et theorie de l'information," p. 178. Presses Universitaires de France, Paris.

MANDELBROT, B. (1957b). Linguistique statistique macroscopique, II. A pamphlet issued by Institut de statistique de l'Université de Paris.